Journal of Computational and Applied Mathematics 234 (2010) 1374–1386

Contents lists available at ScienceDirect

Journal of Computational and Applied Mathematics

journal homepage: www.elsevier.com/locate/cam

Two new conjugate gradient methods based on modified secant equations

Saman Babaie-Kafaki[a], Reza Ghanbari[b], Nezam Mahdavi-Amiri[a,*]

[a] Faculty of Mathematical Sciences, Sharif University of Technology, Tehran, Iran

[b] Department of Mathematics, Ferdowsi University of Mashhad, Mashhad, Iran

ARTICLE INFO

Article history:
Received 23 January 2008
Received in revised form 30 January 2010

Keywords:
Unconstrained optimization
Modified secant equation
Conjugacy condition
Conjugate gradient method
Global convergence

ABSTRACT

Following the approach proposed by Dai and Liao, we introduce two nonlinear conjugate gradient methods for unconstrained optimization problems. One of our proposed methods is based on a modified version of the secant equation proposed by Zhang, Deng and Chen, and Zhang and Xu, and the other is based on the modified BFGS update proposed by Yuan. An interesting feature of our methods is their account of both the gradient and function values. Under proper conditions, we show that one of the proposed methods is globally convergent for general functions and that the other is globally convergent for uniformly convex functions. To enhance the performance of the line search procedure, we also propose a new approach for computing the initial steplength to be used for initiating the procedure. We provide a comparison of implementations of our methods with the efficient conjugate gradient methods proposed by Dai and Liao, and Hestenes and Stiefel. Numerical test results show the efficiency of our proposed methods.

1. Introduction

Conjugate gradient methods have played special roles in solving large scale nonlinear optimization problems. Although conjugate gradient methods are not the fastest or most robust optimization algorithms for nonlinear problems available today, they remain very popular for engineers and mathematicians engaged with solving large problems.

A general conjugate gradient method is designed to solve the following unconstrained optimization problem:

$$\min_{x\in\mathbb{R}^n} f(x), \tag{1.1}$$

where $f:\mathbb{R}^n\to\mathbb{R}$ is a smooth nonlinear function.

Notation 1.1. For a sufficiently smooth function f at x_k, we consider the following notations:

$$f_k = f(x_k), \qquad g_k = \nabla f(x_k), \qquad G_k = \nabla^2 f(x_k). \tag{1.2}$$

The iterative formula of a conjugate gradient method is given by

$$x_{k+1} = x_k + s_k, \quad s_k = \alpha_k d_k, \tag{1.3}$$

* Corresponding author.
E-mail addresses: kafaki@mehr.sharif.ir (S. Babaie-Kafaki), rghanabri@math.um.ac.ir (R. Ghanbari), nezamm@sharif.ir, nezamm@sharif.edu (N. Mahdavi-Amiri).

0377-0427/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.cam.2010.01.052

where α_k is a steplength computed by a line search technique [1], and d_k is the search direction defined by

$$d_k = \begin{cases} -g_k, & k = 1, \\ -g_k + \beta_k d_{k-1}, & k \geq 2, \end{cases} \tag{1.4}$$

where β_k is a scalar.
The following definitions and results are needed (see [2]).

Lemma 1.1. *A quadratic function f, as defined by*

$$f(x) = \frac{1}{2}x^T Hx + b^T x, \tag{1.5}$$

is (strictly) convex if and only if H is a positive semidefinite (definite) matrix.

Definition 1.1. If f is a strictly convex quadratic function and α_k is computed by an exact line search along the direction d_k, i.e.,

$$\alpha_k = \arg\min_{\alpha \geq 0} f(x_k + \alpha d_k), \tag{1.6}$$

then (1.3)–(1.4) is called a linear conjugate gradient method. Otherwise, we have a nonlinear conjugate gradient method.

Remark 1.1. Since an exact line search is not computationally tractable, inexact line search methods have been proposed [1]. Here, we use the so-called strong Wolfe line search conditions [1], i.e., the steplength α_k is computed such that

$$\begin{aligned} &f(x_k + \alpha_k d_k) - f_k \leq \delta \alpha_k g_k^T d_k, \\ &|g(x_k + \alpha_k d_k)^T d_k| \leq -\sigma g_k^T d_k, \end{aligned} \tag{1.7}$$

where $0 < \delta < \sigma < 1$.

Definition 1.2. A twice continuously differentiable function f is said to be uniformly convex on the nonempty open convex set S if and only if there exists $M > 0$ such that

$$(g(x) - g(y))^T (x - y) \geq M\|x - y\|^2, \quad \forall x, y \in S, \tag{1.8}$$

or, equivalently, there exists $r > 0$ such that

$$z^T \nabla^2 f(x) z \geq r\|z\|^2, \quad \forall x \in S, \ \forall z \in \mathbb{R}^n. \tag{1.9}$$

Remark 1.2. In this work, $\|.\|$ stands for Euclidean norm.

Some well-known formulae for β_k in (1.4) are given in [3–8]. The convergence properties of conjugate gradient methods have been studied by many researchers (see [9–16]). Good reviews of the conjugate gradient methods can be found in [17,18].

Although all these methods are equivalent in the linear case, that is, when f is a strictly convex quadratic function and α_k is computed by an exact line search, their behavior for general functions may be quite different. For general functions, Zoutendijk [16] proved the global convergence of the FR method with an exact line search. Although one would be satisfied with global convergence of the FR method, this method performs much worse than the PRP and HS methods. Also, Powell [14] constructed a counter-example and showed that the PRP and HS methods can cycle infinitely without convergence to a solution. This example shows that these two methods have the drawback of not being globally convergent for general functions. Therefore, in the past few years, much effort has been made to find new conjugate gradient methods having not only the global convergence property for general functions but also good numerical performances. These new conjugate gradient methods are based on secant equations (see [19–28]).

As a brief comment, quasi-Newton methods calculate the search direction d_k by solving the following linear system of equations,

$$B_k d_k = -g_k, \tag{1.10}$$

and set $x_{k+1} = x_k + \alpha_k d_k$ at the kth iteration. The matrix B_k is an approximation of G_k in (1.2). The quasi-Newton methods are characterized by the fact that B_k is effectively updated to obtain a new matrix B_{k+1} in the form

$$B_{k+1} = B_k + \Delta B_k,$$

where ΔB_k is a correction matrix [1]. The matrix B_{k+1} is imposed on satisfying a suitable condition which includes the second -order information. The most popular condition is the secant equation,

$$B_{k+1} s_k = y_k, \tag{1.11}$$

where $y_k = g_{k+1} - g_k$. The relation (1.11) is sometimes called the standard secant equation.

Certain kinds of secant equation have been considered to yield better approximations for B_{k+1}. Many researchers have dealt with the extended secant conditions,

$$B_{k+1} s_i = y_i, \quad i = 1, 2, \ldots, k,$$